



Bootstrap Sequential Determination of the Co-integration Rank in VAR Models

Cavaliere, Giuseppe; Rahbek, Anders; Taylor, A. M. Robert

Publication date:
2010

Document version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Cavaliere, G., Rahbek, A., & Taylor, A. M. R. (2010). *Bootstrap Sequential Determination of the Co-integration Rank in VAR Models*. Department of Economics, University of Copenhagen.

Discussion Papers
Department of Economics
University of Copenhagen

No. 10-07

Bootstrap Sequential Determination of
the Co-integration Rank in VAR Models

Giuseppe Cavaliere, Anders Rahbek,
and A. M. Robert Taylor

Øster Farimagsgade 5, Building 26, DK-1353 Copenhagen K., Denmark
Tel.: +45 35 32 30 01 – Fax: +45 35 32 30 00
<http://www.econ.ku.dk>

ISSN: 1601-2461 (online)

Bootstrap Sequential Determination of the Co-integration Rank in VAR Models*

Giuseppe Cavaliere^a, Anders Rahbek^b and A.M.Robert Taylor^c

^a Department of Statistical Sciences, University of Bologna

^b Department of Economics, University of Copenhagen

^c School of Economics and Granger Centre for Time Series Econometrics, University of Nottingham

February 2010

Abstract

Determining the co-integrating rank of a system of variables has become a fundamental aspect of applied research in macroeconomics and finance. It is well-known that standard asymptotic likelihood ratio tests for co-integration rank of Johansen (1996) can be unreliable in small samples with empirical rejection frequencies often very much in excess of the nominal level. As a consequence, bootstrap versions of these tests have been developed. To be useful, however, sequential procedures for determining the co-integrating rank based on these bootstrap tests need to be consistent, in the sense that the probability of selecting a rank smaller than (equal to) the true co-integrating rank will converge to zero (one minus the marginal significance level), as the sample size diverges, for general $I(1)$ processes. No such likelihood-based procedure is currently known to be available. In this paper we fill this gap in the literature by proposing a bootstrap sequential algorithm which we demonstrate delivers consistent co-integration rank estimation for general $I(1)$ processes. Finite sample Monte Carlo simulations show the proposed procedure performs well in practice.

Keywords: Co-integration; trace test; sequential rank determination; i.i.d. bootstrap; wild bootstrap.

J.E.L. Classifications: C30, C32.

*Parts of this paper were written while Cavaliere and Rahbek both visited CREATES whose hospitality is gratefully acknowledged. We are grateful to Bent Nielsen, Niels Haldrup, Søren Johansen and Anders Swensen for many useful discussions on this work. Rahbek is also affiliated with CREATES funded by the Danish National Research Foundation. Correspondence to: Giuseppe Cavaliere, Department of Statistical Sciences, University of Bologna, Bologna, I-40126 Italy. *Authors' E-mail Addresses:* giuseppe.cavaliere@unibo.it (G. Cavaliere), anders.rahbek@econ.ku.dk (A. Rahbek), robert.taylor@nottingham.ac.uk (A.M.R. Taylor)

1 Introduction

Sequential likelihood-based procedures for the determination of the co-integration rank in VAR systems of variables integrated of order 1 [$I(1)$], see Johansen (1996), are extensively used in empirical research. However, it is now well understood that the finite sample properties of these procedures, when based on asymptotic inference, can be quite poor; see, in particular, Johansen (2002) and the references therein. It is also well-known that the bootstrap, when correctly implemented, can be an important device to compute critical values of asymptotic tests in samples of finite size thereby delivering tests with empirical rejection frequencies closer to the nominal level. As a consequence, it is not surprising that there has been an increasing interest in using bootstrap methods in determining the co-integration rank in vector autoregressive models. For co-integrated VAR models with independent and identically distributed (i.i.d.) innovations, see, among others, van Giersbergen (1996), Harris and Judge (1998), Mantalos and Shukur (2001), Swensen (2006, 2009) and Trenkler (2009); for VAR models with potentially heteroskedastic innovations, see Cavaliere *et al.* (2010a, 2010b).

In order to be both operational and efficacious a bootstrap sequential method for determining co-integration rank needs to satisfy the following three requirements: (i) it is applicable to general $I(1)$ systems; (ii) it is asymptotically valid and consistent, where consistency is taken in the usual sense to mean that the probability of selecting a co-integration rank smaller than the true rank converges to zero, while the probability of selecting rank equal to the true co-integrating rank will converge to one minus the chosen (marginal) significance level, as the sample size diverges, and (iii) it improves upon the finite sample performance of the corresponding procedure based on the asymptotic tests. As discussed in Swensen (2009), the sequential algorithm of Swensen (2006, Algorithm 2) does not satisfy these conditions; nor, for exactly the same reason, does Algorithm 1 of Cavaliere *et al.* (2010a, 2010b) which replaces the i.i.d. sub-sampling element of Swensen (2006) with wild bootstrap sub-sampling. In particular, Swensen (2009) demonstrates that the analytic methods used in Swensen (2006) to establish that requirement (ii) above holds for Algorithm 2 are in error and that to do so requires a number of additional conditions to hold on the (unknown) parameters of the data generating process (DGP). These conditions, labelled Assumption 2 in Swensen (2009), are not required by the sequential method based on the asymptotic tests and are violated by a large set of empirically plausible $I(1)$ DGPs, thereby implying that this bootstrap procedure fails requirement (i) above.

Swensen (2009) suggests that two natural responses arise from the problem he highlights. First, he argues that it is important to determine how restrictive the additional conditions are and, second, to formulate a bootstrap algorithm that is consistent even if the additional conditions he lays out do not hold. While there may be some interest in the first issue, it is clear that the key to making progress is to address the second of the responses Swensen (2009) calls for. In this paper we do precisely that and propose an alternative bootstrap sequential algorithm for the determination of the co-integrating rank which meets all of requirements (i)-(iii) above. The proposed algorithm principally differs from Algorithm 2 in Swensen (2006) and Algorithm 1 of Cavaliere *et al.* (2010a, 2010b), by having all parameters used to generate the bootstrap samples re-estimated under the reduced rank being tested at each stage of the procedure. Indeed, this idea was first mentioned in Remark 2 of Swensen (2006, pp.1701-02), although it was not pursued further there other than noting it as a possible alternative to

Algorithm 2 of Swensen (2006). The algorithm we discuss is based on well-known likelihood-based reduced rank estimation and ordinary least squares (OLS), and is therefore entirely straightforward to apply. We establish its asymptotic validity and show that the algorithm can be applied to general $I(1)$ DGPs, without the necessity to satisfy the additional conditions detailed in Assumption 2 of Swensen (2009). The validity of our proposed algorithm is established for both i.i.d. and wild bootstrap re-sampling methods. Furthermore a Monte Carlo simulation study shows that our proposed algorithm also improves on the finite sample properties of existing procedures.

The remainder of the paper is organised as follows. In section 2 we provide a brief summary of the model and standard asymptotic theory for sequential rank determination. In section 3 we present the proposed sequential bootstrap algorithm, demonstrating how this differs from Algorithm 2 of Swensen (2006) and Algorithm 1 of Cavaliere *et al.* (2010a, 2010b). The asymptotic properties of our proposed procedure are investigated in section 4, while a small Monte Carlo study of its finite sample behaviour is given in section 5. Section 6 concludes. Mathematical proofs are contained in the Appendix.

In the following ‘ \xrightarrow{w} ’ denotes weak convergence, ‘ \xrightarrow{p} ’ convergence in probability, and ‘ \xrightarrow{w}_p ’ weak convergence in probability (Giné and Zinn, 1990; Hansen, 1996), in each case as the sample size diverges to positive infinity; ‘ $\mathbb{I}(\cdot)$ ’ denotes the indicator function; ‘ $x := y$ ’ indicates that x is defined by y ; $\lfloor \cdot \rfloor$ denotes the integer part of its argument; I_k denotes the $k \times k$ identity matrix and $0_{j \times k}$ the $j \times k$ matrix of zeroes; the space spanned by the columns of any $m \times n$ matrix A is denoted as $\text{col}(a)$; if a is of full column rank $n < m$, then $\bar{a} := a(a'a)^{-1}$ and a_\perp is an $m \times (m - n)$ full column rank matrix satisfying $a'_\perp a = 0$; for any square matrix, a , $|a|$ is used to denote its determinant, $\|a\|$ the norm $\|a\|^2 := \text{tr}\{a'a\}$ (where $\text{tr}\{a\}$ denotes the trace of a) and $\rho(a)$ its spectral radius (that is, the maximal modulus of the eigenvalues of a); for any vector, x , $\|x\|$ denotes the usual Euclidean norm, $\|x\| := (x'x)^{1/2}$. Finally, P^* denotes the bootstrap probability measure, i.e. conditional on the original sample; similarly, E^* denotes expectation under P^* .

2 The Co-integration Model and Asymptotic Test Procedures

We consider the usual $\text{VAR}(k)$ model in error correction format:

$$\Delta X_t = \Pi X_{t-1} + \Psi U_t + \mu D_t + \varepsilon_t, \quad t = 1, \dots, T \quad (1)$$

where $X_t := (X_{1t}, \dots, X_{pt})'$ and the innovations, $\varepsilon_t := (\varepsilon_{1t}, \dots, \varepsilon_{pt})'$, are both $p \times 1$, and $U_t := (\Delta X'_{t-1}, \dots, \Delta X'_{t-k+1})'$ is $p(k-1) \times 1$, $\Psi := (\Gamma_1, \dots, \Gamma_{k-1})$, where $\{\Gamma_i\}_{i=1}^{k-1}$ are $p \times p$ lag coefficient matrices. The impact matrix $\Pi := \alpha\beta'$, where α and β are $p \times r$, $r \leq p$, matrices, with the usual convention that $\alpha\beta'$ is the $p \times p$ matrix of zeroes when $r = 0$ (no co-integration). The initial values, $\mathbb{X}_0 := (X'_0, \dots, X'_{-k+1})'$, are taken to be fixed. The model, which will often be referred to as model $H(r)$ in what follows, may be written in the compact form,

$$Z_{0t} = \alpha\beta^\# Z_{1t} + \Psi^\# Z_{2t} + \varepsilon_t \quad (2)$$

with $Z_{0t} := \Delta X_t$, and the remaining terms defined according to the following three leading cases for the deterministic term, μD_t (see, e.g., Johansen, 1996, p.81):

- (i) $\mu D_t = 0$ in (1), which implies that $Z_{1t} := X_{t-1}$, $Z_{2t} := U_t$, $\beta^\# = \beta$ and $\Psi^\# = \Psi$ (no deterministic components);
- (ii) $\mu D_t = \mu_1 = \alpha \rho'$ in (1), which implies that $Z_{1t} := (X'_{t-1}, 1)'$, $Z_{2t} := U_t$, $\beta^\# = (\beta', \rho')'$ and $\Psi^\# = \Psi$ (restricted constant);
- (iii) $\mu D_t = \mu_1 + \mu_2 t$ with $\mu_2 = \alpha \rho'$ in (1), which implies that $Z_{1t} := (X'_{t-1}, t)'$, $Z_{2t} := (U'_t, 1)'$, $\beta^\# = (\beta', \rho')'$ and $\Psi^\# = (\Psi, \mu_1)$ (restricted linear trend).

Throughout the paper the process in (1) is assumed to satisfy the following conditions. First, as is standard (see, e.g., Engle and Granger, 1987), the autoregressive coefficients in (1) are assumed to satisfy the following set of assumptions.

Assumption 1: (a) *all the characteristic roots associated with (1); that is of $A(z) := (1 - z)I_p - \alpha\beta'z - \Gamma_1z(1 - z) - \dots - \Gamma_{k-1}z^{k-1}(1 - z) = 0$, are outside the unit circle or equal to 1;* (b) α and β *have full column rank r ;* (c) $\det(\alpha'_\perp \Gamma \beta_\perp) \neq 0$, with $\Gamma := I_p - \Gamma_1 - \dots - \Gamma_{k-1}$.

Second, as is routinely done in this literature (see, e.g., Johansen, 1996, Swensen, 2006), we assume that the innovations are i.i.d. While this assumption is convenient for expositional purposes it is also possible to derive corresponding results under (either conditionally or unconditionally) heteroskedastic innovations; see Cavaliere *et al.* (2010a, 2010b). We will briefly discuss these generalisations later.

Assumption 2: *The innovations $\{\varepsilon_t\}$ form an i.i.d. sequence with $E(\varepsilon_t) = 0$ and $E(\varepsilon_t \varepsilon'_t) = \Omega$, with Ω positive definite, and (ii) $E\|\varepsilon_t\|^4 \leq K < \infty$.*

In the following, a VAR model satisfying (1) under Assumptions 1 and 2 will be said to satisfy the 'I(1, r)' conditions. Moreover, we let r_0 denote the true (unknown) rank of Π .

Now, consider the problem of determining the co-integration rank. As is standard, let $M_{ij} := T^{-1} \sum_{t=1}^T Z_{it} Z'_{jt}$, $i, j = 0, 1, 2$, with Z_{it} defined as in (2), and let $S_{ij} := M_{ij} - M_{i2} M_{22}^{-1} M_{2j}$, $i, j = 0, 1$. Then, as shown in Johansen (1996), the LR test for $H(r)$ vs $H(p)$, $r = 0, \dots, p-1$, rejects for large values of the trace statistic,

$$Q_{r,T} := -T \sum_{i=r+1}^p \log(1 - \hat{\lambda}_i) \quad (3)$$

where $\hat{\lambda}_1 > \dots > \hat{\lambda}_p$ are the largest p solutions to the eigenvalue problem,

$$|\lambda S_{11} - S_{10} S_{00}^{-1} S_{01}| = 0. \quad (4)$$

Under Assumptions 1 and 2, Johansen (1996) establishes that, as $T \rightarrow \infty$, so

$$Q_{r,T} \rightarrow +\infty, \text{ for } r = 0, \dots, r_0 - 1 \quad (5)$$

while, for the true co-integrating rank,

$$Q_{r_0, T} \xrightarrow{w} \text{tr}(\mathcal{Q}_{r_0, \infty}) \quad (6)$$

where, for $0 \leq r \leq p-1$,

$$\mathcal{Q}_{r, \infty} := \int_0^1 dB_{p-r}(u) F_{p-r}(u)' \left(\int_0^1 F_{p-r}(u) F_{p-r}(u)' du \right)^{-1} \int_0^1 F_{p-r}(u) dB_{p-r}(u)' \quad (7)$$

B_{p-r} denoting a $(p-r)$ -variate standard Brownian motion and where F_{p-r} is a function of B_{p-r} and the included deterministic term(s). More specifically, using the notation $a|b := a(\cdot) - \int a(s)b(s)'ds (\int b(s)b(s)'ds)^{-1}b(\cdot)$ to denote the projection residuals of a onto b , then: (i) if $\mu D_t = 0$ in (1) and no deterministics are included in the estimation, then $F_{p-r} := B_{p-r}$; (ii) if $\mu D_t = \alpha \rho'$ in (1) and a restricted constant is included in the estimation, then $F_{p-r} := (B_{p-r}', 1)'$; (iii) if $\mu D_t = \mu_1 + \alpha \rho' t$ in (1) and a restricted linear trend is included in the estimation, then $F_{p-r} := (B_{p-r}', u|1)'$.

For each of $r = 0, \dots, p-1$, let G_r^∞ denote the cumulative density function [cdf] of $\text{tr}(\mathcal{Q}_{r_0, \infty})$. Starting with $r = 0$, the sequential procedure of Johansen (1996) involves testing in turn $H(r)$ against $H(p)$ for $r = 0, \dots, p-1$, until, for a given value of r , the asymptotic p -value, $p_{r, T}^\infty := 1 - G_r^\infty(Q_{r, T})$ exceeds a chosen (marginal) significance level, say η . The estimated co-integration rank, say \hat{r} , is then the lowest value of r such that the corresponding asymptotic p -value is above η , with \hat{r} set to p if $p_{r, T}^\infty \leq \eta$ for all $r = 0, 1, \dots, p-1$. The results in (5) and (6) imply that the sequential LR procedure is consistent in the sense that $\lim_{T \rightarrow \infty} P(\hat{r} = r) = 0$, for $r < r_0$, and $\lim_{T \rightarrow \infty} P(\hat{r} = r_0) = 1 - \eta \cdot \mathbb{I}(r_0 < p)$; see Johansen (1996, Chapter 12).

Remark 2.1. The preceding discussion extends to the so-called maximum eigenvalue test; that is, the LR test based for $H(r)$ vs $H(r+1)$. As is well known, this test rejects for large values of the statistic $Q_{r, \max} := -2(\ell(r) - \ell(r+1)) = -T \log(1 - \hat{\lambda}_{r+1})$; see, for example, Equation (6.19) of Johansen (1996). The null asymptotic distribution of $Q_{r, \max}$ corresponds to the distribution of the maximum eigenvalue of the real symmetric random matrix $\mathcal{Q}_{r, \infty}$, while a sequential approach based on $Q_{r, \max}$, $r = 0, \dots, p-1$, will share the same consistency properties given above for the corresponding sequential procedure based on the trace statistics; see Paruolo (2001).

Remark 2.2. Cavaliere *et al.* (2010a) demonstrate that the large sample results given in this section remain valid when the innovation process $\{\varepsilon_t\}$ in (1) is a vector martingale difference sequence with respect to the filtration \mathcal{F}_t , where $\mathcal{F}_{t-1} \subseteq \mathcal{F}_t$ for $t = \dots, -1, 0, 1, 2, \dots$, which satisfies both a fourth moment condition (see Assumption 2(ii)) and the *global homoskedasticity* condition, $T^{-1} \sum_{t=1}^T \mathbb{E}(\varepsilon_t \varepsilon_t' | \mathcal{F}_{t-1}) \xrightarrow{P} \Omega$, with Ω positive definite. This assumption therefore allows for certain forms of conditional heteroskedasticity in the innovations. Unconditional heteroskedasticity, as considered by Cavaliere *et al.* (2010b), can, however, alter the large sample results given in this section. Precisely, using the decomposition $\varepsilon_t = \sigma_t z_t$ with z_t a p -variate i.i.d. process and σ_t a non-stochastic matrix satisfying: (i) $\sigma_t := \sigma(t/T)$ for all $t = 1, \dots, T$, where $\sigma(\cdot)$ lies in the space of $p \times p$ matrices of càdlàg functions on $[0, 1]$, and (ii) $\Sigma(u) := \sigma(u) \sigma(u)'$ is assumed positive definite for all $u \in [0, 1]$, Cavaliere *et al.* (2010b)

show that although the result in (5) remains valid, the result in (6) is no longer appropriate, but rather that, in the case of the trace statistic, $Q_{r_0,T} \xrightarrow{w} \text{tr}(Q_{r_0,\infty}^\sigma)$, where $Q_{r_0,\infty}^\sigma$ is as given by the right member of (7) but with the $(p-r)$ -variate stochastic volatility process $\tilde{M} := (\alpha'_\perp \bar{\Sigma} \alpha'_\perp)^{-1/2} \alpha'_\perp \int_0^1 \sigma(s) dB_p(s)$ replacing B_{p-r} throughout (including in the definition of F_{p-r}), where $\bar{\Sigma} := \int_0^1 \Sigma(s) ds$, so that the standard asymptotic test will therefore be incorrectly sized, even in the limit. The same comments hold (the relevant limiting distribution now being that of the maximum eigenvalue of $Q_{r,\infty}^\sigma$) for the maximum eigenvalue statistic.

3 The Bootstrap Sequential Algorithm

Bootstrap analogues of the $Q_{r,T}$ trace statistic (and corresponding maximum eigenvalue statistic) from section 2 have been developed in the literature. This is done for the i.i.d. bootstrap in Swensen (2006, Algorithm 1) and for the wild bootstrap (see, *inter alia*, Wu, 1986, Liu, 1988 and Mammen, 1992) in Algorithm 1 of Cavaliere *et al.* (2010a,2010b). Swensen (2006) established that, under Assumptions 1 and 2, the i.i.d. bootstrap analogue of the $Q_{r_0,T}$ trace statistic replicates that statistic's first order limiting null distribution when the true rank is r_0 . Cavaliere *et al.* (2010a) show that this result also holds for both the i.i.d. and wild bootstrap statistics under the martingale difference conditions laid out in Remark 2.2, while Cavaliere *et al.* (2010b) prove that in the case of unconditional heteroskedasticity as in Remark 2.2, the wild (but not the i.i.d.) bootstrap statistic attains the same first order (non-pivotal) limiting distribution, $\text{tr}(Q_{r_0,\infty}^\sigma)$, as $Q_{r_0,T}$ under rank r_0 , such that asymptotically correctly sized inference can be obtained using the wild bootstrap.

However, and as argued in Swensen (2006,p.1702), the most important use of the likelihood ratio test for co-integration rank is as part of a sequential procedure to determine the co-integration rank. As Swensen (2006,p.1700) notes, "This problem is more intriguing and presents some new aspects that are nonstandard in a bootstrap context, because we have to do the resampling for different values of the rank of the estimated reduced rank matrix. The dimension of the cointegration space in the generated observations will therefore not correspond to the true cointegration rank, but to the imposed rank ...". As is done in Algorithm 2 of Swensen (2006,pp.1702-3) and in Algorithm 1 of Cavaliere *et al.* (2010a,2010b) this is implemented by estimating α , β (and ρ in the case of deterministic terms) under $H(r)$, and then using these estimated values in the re-sampling recursion equation (see, for example, Equation (4) in step (iv) of Algorithm 2 of Swensen, 2006). These algorithms, however, differ from the algorithm we now outline in that the remaining parameters from (1) are estimated unrestrictedly, i.e. under $H(p)$.

At each stage of our proposed sequential algorithm, the bootstrap trace statistic is calculated from re-sampled data which are constructed using the (Gaussian) (quasi-) likelihood-based estimates [QMLE] under $H(r)$: notice that this model will be misspecified unless $r = r_0$. Denote the corresponding estimators by $\hat{\beta}^{(r)}$, $\hat{\alpha}^{(r)}$, $\hat{\Psi}^{(r)} := (\hat{\Gamma}_1^{(r)}, \dots, \hat{\Gamma}_{k-1}^{(r)})$, $\hat{\Omega}^{(r)}$ and $\hat{\mu}^{(r)}$, which are obtained in the usual manner. That is, in the case of no deterministic terms, with the ordered eigenvalues, $\hat{\lambda}_1 > \dots > \hat{\lambda}_p$, obtained as laid out in section 2, let $\hat{v} := (\hat{v}_1, \hat{v}_2, \dots, \hat{v}_p)$ denote the corresponding eigenvectors, which satisfy

$$\hat{v}' S_{11} \hat{v} = I_p, \quad \hat{v}' S_{10} S_{00}^{-1} S_{01} \hat{v} = \hat{\Lambda}_p := \text{diag}(\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_p). \quad (8)$$

For each $r = 0, 1, \dots, p-1$, the (uniquely defined) QMLE, $\hat{\beta}^{(r)}$, is then given by

$$\hat{\beta}^{(r)} := \hat{v} K_p^{(r)},$$

where $K_p^{(r)} := (I_r, 0_{r \times (p-r)})'$, is a selection matrix indexed by r and p . When deterministic terms are included, $\hat{\beta}^{\#(r)} := (\hat{\beta}^{(r)'}, \hat{\rho}^{(r)'})' = \hat{v} K_{p^\#}^{(r)}$ as in this case the eigenvectors in \hat{v} are of dimension $p^\# = p + 1$. After $\hat{\beta}^{(r)}$, or $\hat{\beta}^{\#(r)}$, is computed, the remaining estimators are obtained by simple OLS regression, as detailed in Johansen (1996).

Using these restricted estimators, our proposed bootstrap sequential procedure is then as follows.

Algorithm 1 *Starting from $r = 0$, perform the following steps:*

- (i) *Estimate model (1) under rank r using Gaussian QMLE as outlined above and denote the resulting residuals by $\hat{\varepsilon}_{r,t}$.*
- (ii) *Generate T bootstrap errors $\varepsilon_{r,t}^*$ using the re-centred residuals¹, $\hat{\varepsilon}_{r,t}^c := \hat{\varepsilon}_{r,t} - T^{-1} \sum_{i=1}^T \hat{\varepsilon}_{r,i}$, for either: (a) the i.i.d. bootstrap, such that $\varepsilon_{r,t}^* := \hat{\varepsilon}_{r,\mathcal{U}_t}^c$, where \mathcal{U}_t , $t = 1, \dots, T$ is an i.i.d. sequence of discrete uniform distributions on $\{1, 2, \dots, T\}$, or (b) the wild bootstrap, where for each $t = 1, \dots, T$, $\varepsilon_{r,t}^* := \hat{\varepsilon}_{r,t}^c w_t$, where w_t , $t = 1, \dots, T$, is an i.i.d. $N(0, 1)$ sequence.*
- (iii) *Construct the bootstrap sample recursively from*

$$\Delta X_{r,t}^* := \hat{\alpha}^{(r)} \hat{\beta}^{(r)'} X_{r,t-1}^* + \hat{\Gamma}_1^{(r)} \Delta X_{r,t-1}^* + \dots + \hat{\Gamma}_{k-1}^{(r)} \Delta X_{r,t-k+1}^* + \varepsilon_{r,t}^*, \quad t = 1, \dots, T, \quad (9)$$

initialised at $X_{-k+1}^ = \dots = X_0^* = 0$;*

- (iv) *Using the bootstrap sample, $\{X_{r,t}^*\}$, and denoting by $\hat{\lambda}_1^* > \dots > \hat{\lambda}_p^*$ the ordered solutions to the bootstrap analogue of the eigenvalue problem in (4), compute the LR test statistic*

$$Q_{r,T}^* := -T \sum_{i=r+1}^p \log(1 - \hat{\lambda}_i^*)$$

for the null of rank r against rank p , along with the corresponding p -value $p_{r,T}^ := 1 - G_{r,T}^*(Q_{r,T}^*)$, where $G_{r,T}^*(\cdot)$ denotes the conditional (on the original data) cdf of $Q_{r,T}^*$;*

- (v) *If $p_{r,T}^*$ exceeds the significance level, η , set $\hat{r} = r$, otherwise repeat steps (i)–(iv) testing the null of rank $(r+1)$ against rank p if $r+1 < p$, or set $\hat{r} = p$ if $r+1 = p$.*

¹Observe that re-centring is not needed in the case of a restricted trend as $\sum_{i=1}^T \hat{\varepsilon}_{r,i} = 0$ in this case.

Remark 3.1. Algorithm 1 differs from the corresponding algorithm in Swensen (2006, Algorithm 2) in a number of respects. Firstly, although Algorithm 2 of Swensen (2006) also suggests estimating α and β under rank r , in contrast to our proposed approach in Algorithm 1, Algorithm 2 of Swensen (2006) estimates Ψ unrestrictedly; i.e., under rank p , we denote this estimator by $\hat{\Psi}^{(p)} := (\hat{\Gamma}_1^{(p)}, \dots, \hat{\Gamma}_{k-1}^{(p)})$. Similarly, in step (ii) of Algorithm 2, Swensen (2006) uses residuals from the unrestricted model (i.e., those obtained under rank p) which are therefore the same at each stage of the algorithm, while our Algorithm 1 employs the restricted residuals at each stage. As we will demonstrate in section 4, using the restricted estimators ensures that the re-sampling recursion equation (9) always delivers an $I(1, r)$ system in the limit with $p - r$ common stochastic $I(1)$ trends, and $r, r \leq r_0$, co-integrating vectors. As recognised in Swensen (2009), this is not guaranteed under Algorithm 2 of Swensen (2006), nor indeed Algorithm 1 of Cavaliere *et al.* (2010a, 2010b), since when Ψ is estimated unrestrictedly, the bootstrap samples will not be asymptotically $I(1, r)$ unless Assumption 2 of Swensen (2009) is met; see also section 5 below.

Remark 3.2. Algorithm 1 differs from Algorithm 2 of Swensen (2006) in a further way. In step (iii) of Algorithm 2 of Swensen (2006) one must check, for each value of r tested, that the roots of the equation $|\hat{A}^{(r)}(z)| = 0$ are either one or outside the unit circle, where

$$\hat{A}^{(r)}(z) := (1 - z)I_p - \hat{\alpha}^{(r)}\hat{\beta}^{(r)'}z - \hat{\Gamma}_1^{(p)}(1 - z)z - \dots - \hat{\Gamma}_{k-1}^{(p)}(1 - z)z^{k-1},$$

together with the requirement that $|\hat{\alpha}_\perp^{(r)'}\hat{\Gamma}_\perp^{(p)}\hat{\beta}_\perp^{(r)}| \neq 0$, where $\hat{\Gamma}^{(p)} := I_p - \hat{\Gamma}_1^{(p)} - \dots - \hat{\Gamma}_{k-1}^{(p)}$, and $\hat{\Gamma}_j^{(p)}, j = 1, \dots, k - 1$, are the unrestricted estimates defined in Remark 3.1. It can be shown that step (iii) will fail for the relevant value(s) of r with probability tending to one as the sample size increases if Assumption 2 of Swensen (2009) does not hold for those value(s) of r . In cases where step (iii) fails, Swensen (2006, p.1701) argues that “... another more appropriate recursive scheme that reflects the properties of the observed data should be used.” which is just what Algorithm 1 of this paper is designed to provide. It is also, therefore, not necessary to check this condition when using the procedure outlined in Algorithm 1 of this paper.

Remark 3.3. Notice also that, due to the (exact) invariance of $Q_{r,T}$ with respect to μ , we need not add an estimate of the deterministic component, μD_t , to the right member of (9) as is done in, for example, Algorithm 2 of Swensen (2006). Moreover, since $Q_{r,T}$ is similar with respect to the initial values (exact similar under cases (ii) and (iii) for μD_t given in section 2 and asymptotically similar under case (i)) we may set these to zero in our recursive scheme. As an alternative to (9) one could use the recursion

$$\Delta X_{r,t}^* := \hat{\alpha}^{(r)}\hat{\beta}^{(r)'}X_{r,t-1}^* + \hat{\Gamma}_1^{(r)}\Delta X_{r,t-1}^* + \dots + \hat{\Gamma}_{k-1}^{(r)}\Delta X_{r,t-k+1}^* + \hat{\mu}^{(r)}D_t + \varepsilon_{r,t}^*, \quad t = 1, \dots, T$$

with initial values, $X_t^* := X_t, t = -k + 1, \dots, 0$. In unreported Monte Carlo simulations we found no discernible differences between the finite sample properties of these two approaches and so we have adopted the simpler of the two.

Remark 3.4. In practice, the cdf $G_{r,T}^*(\cdot)$ required in Step (iv) of Algorithm 1 will not be known, but can be approximated in the usual way through numerical simulation; cf. Hansen (1996) and Andrews and Buchinsky (2000). Taking the case of the wild bootstrap to illustrate, this is achieved by generating B (conditionally) independent bootstrap statistics, $Q_{r,b,T}^*$, $b = 1, \dots, B$, computed as above but recursively from

$$\Delta X_{r,b,t}^* := \hat{\alpha}^{(r)} \hat{\beta}^{(r)'} X_{r,b,t-1}^* + \hat{\Gamma}_1^{(r)} \Delta X_{r,b,t-1}^* + \dots + \hat{\Gamma}_{k-1}^{(r)} \Delta X_{r,b,t-k+1}^* + \varepsilon_{r,b,t}^*, \quad t = 1, \dots, T,$$

with initial values $X_{r,b,-k+1}^* = \dots = X_{r,b,0}^* = 0$ and with $\{\{w_{r,b,t}\}_{t=1}^T\}_{b=1}^B$ a doubly independent $N(0, 1)$ sequence. The simulated bootstrap p -value is then computed as $\tilde{p}_{r,T}^* := B^{-1} \sum_{b=1}^B \mathbb{I}(Q_{r,b,T}^* > Q_{r,T})$, and is such that $\tilde{p}_{r,T}^* \xrightarrow{a.s.} p_{r,T}^*$ as $B \rightarrow \infty$. Note that an asymptotic standard error for $\tilde{p}_{r,T}^*$ is given by $(\tilde{p}_{r,T}^*(1 - \tilde{p}_{r,T}^*)/B)^{1/2}$; cf. Hansen (1996, p.419).

4 Asymptotic Analysis

In this section prove that Algorithm 1 is consistent for any DGP which satisfies the I(1) conditions stated in Swensen (2006) which consist of the I(1, r) conditions stated previously in section 2, supplemented by Assumption 3 below; that is, our proposed algorithm does not require the additional Assumption 2 of Swensen (2009) to hold for consistency.

By Johansen (1996), under the I(1, r) conditions, the largest r_0 ordered sample eigenvalues, $\hat{\lambda}_1 > \hat{\lambda}_2 > \dots > \hat{\lambda}_{r_0}$ of (4) converge, as $T \rightarrow \infty$, to the corresponding r_0 population eigenvalues which solve (A.1); $\lambda_1, \lambda_2, \dots, \lambda_{r_0}$ say. For the asymptotic analysis, Swensen (2006, Lemma 3) makes the standard assumption that these population eigenvalues are distinct, and we state this here as Assumption 3. This implies that both the population eigenvectors and eigenvalues are continuous functions of the population parameters.

Assumption 3: *The limiting non-zero roots of (4) are distinct; that is, $\lambda_1 > \lambda_2 > \dots > \lambda_{r_0} > 0$.*

Our first lemma concerns the limiting behaviour of the Gaussian QMLE when an incorrect rank $r < r_0$ is imposed. This estimator is used to generate the bootstrap samples, see steps (i) and (iii) of Algorithm 1 of section 3, and we show that in the limit this satisfies the I(1, r) conditions.

Lemma 1 *Let $\{X_t\}$ be generated as in (1) under Assumptions 1–3. Furthermore, let $\hat{\theta}^{(r)} := \{\hat{\alpha}^{(r)}, \hat{\beta}^{(r)}, \hat{\Psi}^{(r)}, \hat{\mu}^{(r)}, \hat{\Omega}^{(r)}\}$ denote the QMLE in (1) under $H(r)$. Then, for any $r < r_0$ and as $T \rightarrow \infty$, $\hat{\theta}^{(r)} \xrightarrow{p} \theta_0^{(r)}$, with $\theta_0^{(r)} := \{\alpha_0^{(r)}, \beta_0^{(r)}, \Psi_0^{(r)}, \mu_0^{(r)}, \Omega_0^{(r)}\}$ defined in the appendix. Moreover, the characteristic equation*

$$A^{(r)}(z) = I_p(1 - z) - \alpha_0^{(r)} \beta_0^{(r)'} z - \Psi_0^{(r)}(1 - z)(z, \dots, z^{k-1})' = 0 \quad (10)$$

satisfies the I(1, r) conditions.

Remark 4.1. Taken together, the results in Lemma 1 imply that, as T increases, the estimates of $\alpha_0^{(r)}, \beta_0^{(r)}, \Psi_0^{(r)} = (\Gamma_{10}^{(r)}, \dots, \Gamma_{k-1}^{(r)})$ satisfy the I(1, r) conditions, even if r is lower than

the true rank r_0 . In particular, (10) implies that $\alpha_{0\perp}^{(r)'} \left(I - \sum_{i=1}^{k-1} \Gamma_{0i}^{(r)} \right) \beta_{0\perp}^{(r)} = \alpha_{0\perp}^{(r)'} \Gamma_0^{(r)} \beta_{0\perp}^{(r)}$ is of full rank $p - r > p - r_0$. Another way of stating this result is simply that the stochastic difference equation

$$\Delta X_t^* = \alpha_0^{(r)} \beta_0^{(r)'} X_{t-1}^* + \Psi_0^{(r)} U_t^* + \varepsilon_t^*,$$

with $\varepsilon_t^* \sim i.i.d.(0, \Omega^{(r)})$ generates an $I(1, r)$ system with $p - r$ common stochastic $I(1)$ trends, and $r < r_0$ co-integrating vectors. This is the key result needed to prove the validity of our sequential algorithm.

The implications of Lemma 1 for the bootstrap DGP are collected in the following representation result, which establishes that for $r \leq r_0$ the bootstrap sample always satisfies the $I(1, r)$ conditions as the sample size increases. This proposition holds irrespective of whether an i.i.d. or wild bootstrapping re-sampling design is used.

Proposition 1 *Let the bootstrap sample be generated as in Algorithm 1 for any $r \leq r_0$. Then, under the conditions of Lemma 1, and if $r \leq r_0$, it holds that*

$$X_{r,t}^* = \hat{C}^{(r)} \sum_{i=1}^t \varepsilon_{r,t}^* + S_{r,t} T^{1/2}$$

where $P^* (\max_{t=1, \dots, T} |S_{r,t}| > \epsilon) \xrightarrow{p} 0$ and $\hat{C}^{(r)} \xrightarrow{p} C_0^{(r)} := \beta_{0\perp}^{(r)} \left(\alpha_{0\perp}^{(r)'} \Gamma_0^{(r)} \beta_{0\perp}^{(r)} \right)^{-1} \alpha_{0\perp}^{(r)'}$, $C_0^{(r)}$ being of rank $p - r$. Moreover,

$$T^{-1/2} X_{r, [T]}^* \xrightarrow{w_p} C_0^{(r)} W(\cdot)$$

where W is a p -dimensional Brownian motion with covariance matrix $\Omega_0^{(r)}$.

Remark 4.2. Proposition 1 shows that, for any $r \leq r_0$, the bootstrap sample (asymptotically) behave as an $I(1, r)$ process. This result does not hold for other extant bootstrap algorithms, such as Algorithm 2 of Swensen (2006), since the latter may generate non- $I(1)$ samples, even asymptotically.

Remark 4.3. The proof of Proposition 1 exploits the fact that, for any rank $r \leq r_0$, the bootstrap sample $\Delta X_{r,t}^* := \hat{\alpha}^{(r)} \hat{\beta}^{(r)'} X_{r,t-1}^* + \hat{\Psi}^{(r)} U_t^* + \varepsilon_{r,t}^*$, approximately behaves, as T increases, as $\Delta X_t^* := \alpha_0^{(r)} \beta_0^{(r)'} X_{r,t-1}^* + \Psi_0^{(r)} U_t^* + \varepsilon_{r,t}^*$. As established in Lemma 1, the characteristic polynomial associated with the latter process satisfies the $I(1, r)$ conditions. This property implies that, asymptotically, the bootstrap sample is $I(1, r)$ with r co-integrating relations and $p - r$ unit roots.

A direct consequence of Lemma 1 and Proposition 1 is the following proposition, where it is shown that the bootstrap trace statistic is asymptotically distributed as $tr(\mathcal{Q}_{r,\infty})$ for any $r \leq r_0$.

Proposition 2 *Let the sequence of bootstrap statistics $Q_{0,T}^*, Q_{1,T}^*, \dots, Q_{r,T}^*$, $r \leq r_0$ be generated as in Algorithm 1. Then, under the conditions of Lemma 1, and for any $r \leq r_0$,*

$$Q_{r,T}^* \xrightarrow{w}_p \text{tr}(\mathcal{Q}_{r,\infty}). \quad (11)$$

Consequently, $p_{r,T}^ := 1 - G_{r,T}^*(Q_{r,T}) \xrightarrow{p} 0$ for all r such that $0 \leq r < r_0$, and $p_{r_0,T}^* \xrightarrow{w} U[0, 1]$.*

We are now ready to establish the overall consistency of the bootstrap sequential procedure embodied in Algorithm 1.

Theorem 1 *Let \hat{r} denote the estimator of the co-integration rank as obtained in Algorithm 1. Then, under the conditions of Lemma 1,*

$$\begin{aligned} \lim_{T \rightarrow \infty} P(\hat{r} = r) &= 0 \text{ for all } r = 0, 1, \dots, r_0 - 1 \\ \lim_{T \rightarrow \infty} P(\hat{r} = r_0) &= 1 - \eta \cdot \mathbb{I}(r_0 < p) \\ \lim_{T \rightarrow \infty} \sup_{r \in \{r_0+1, \dots, p\}} P(\hat{r} = r) &\leq \eta. \end{aligned}$$

Remark 4.4. The results in Theorem 1 establish that the bootstrap sequential procedure outlined in Algorithm 1 replicates the first order asymptotic properties of the corresponding procedure based on the asymptotic tests, outlined in Section 2. This is achieved without the need for the additional conditions in Assumption 2 of Swensen (2009) to hold.

Remark 4.5. Given the preceding results it is straightforward but tedious to show that the large sample results given in this section remain valid under the martingale difference conditions laid out in Remark 2.2. The stated results in Theorem 1 also apply for wild bootstrap (but not i.i.d.) re-sampling under unconditional heteroskedasticity of the form considered in Cavaliere et al. (2010b), as laid out in Remark 2.2. Moreover, all of the conclusions drawn in this section for procedures based on trace statistics trivially also apply to the corresponding procedures based on maximal eigenvalue statistics; cf. Remarks 2.1 and 2.2.

5 Numerical Results

In this section we use Monte Carlo simulation methods to investigate the finite sample performance of the sequential procedure proposed in Algorithm 1 of this paper, comparing its performance, for both i.i.d. and wild bootstrap re-sampling schemes, with the asymptotic procedure of Johansen (1996). Some comparison is also made with Algorithm 2 of Swensen (2006) and the wild bootstrap version of that from Cavaliere *et al* (2010a, 2010b).

The simulation model we consider in this section is the co-integrated VAR(2) process of dimension $p = 4$,

$$\Delta X_t = \alpha \beta' X_{t-1} + \Gamma_1 \Delta X_{t-1} + \varepsilon_t, \quad t = 1, \dots, T \quad (12)$$

initialised at zero. Results are reported for samples of size $T = 100$ and $T = 250$. Within the context of (12) we will consider the following four models for the innovation process ε_t , with parameters set as in Gonçalves and Kilian (2004) and Cavaliere *et al.* (2010b):

- **Model IID:** the innovations are Gaussian i.i.d.; $\varepsilon_t \sim \text{i.i.d. } N(0, I_4)$.
- **Model GARCH:** the innovation ε_{it} follows a standard stable GARCH(1,1) process driven by standard normal innovations of the form $\varepsilon_{it} = h_{it}^{1/2} v_{it}$, $i = 1, \dots, p$, where v_{it} is i.i.d. $N(0, 1)$, independent across i , and $h_{it} = \omega + d_0 \varepsilon_{it-1}^2 + d_1 h_{it-1}$, $t = 0, \dots, T$. Results are reported for $(d_0, d_1) = (0.20, 0.79)$.
- **Model SV:** the innovation ε_{it} follows the stable stochastic volatility process, $\varepsilon_{it} = v_{it} \exp(h_{it})$, $h_{it} = \lambda h_{it-1} + 0.5 \xi_{it}$, with $(\xi_{it}, v_{it}) \sim \text{i.i.d. } N(0, \text{diag}(\sigma_\xi^2, 1))$, independent across $i = 1, \dots, p$. Results are reported for $(\lambda, \sigma_\xi) = (0.951, 0.314)$.
- **Model NSV:** the innovation process displays a one-time break in its unconditional variance, such that $\varepsilon_t = \sigma_t z_t$, where $z_t \sim \text{i.i.d. } N(0, I_4)$ and $\sigma_t = 1 + \varpi \mathbb{I}(t > \lfloor \tau T \rfloor)$, where $\mathbb{I}(\cdot)$ is the usual indicator function. Results are reported for $\varpi = 5$ and $\tau = 0.9$.

Models GARCH and SV allow for conditional heteroskedasticity, which along with Model IID, both satisfy the global homoskedasticity condition of Remark 2.2. In contrast, the shocks from Model NSV are unconditionally heteroskedastic.

In order to isolate clearly the failure or otherwise of the extra conditions given in Assumption 2 of Swensen (2009) through a single parameter, we will focus attention on the specific form of (12) where $\beta := (1, 0, 0, 0)'$, $\alpha := (a_1, 0, 0, 0)'$, and

$$\Gamma_1 := \begin{bmatrix} \gamma & \delta & 0 & 0 \\ \delta & \gamma & 0 & 0 \\ 0 & 0 & \gamma & 0 \\ 0 & 0 & 0 & \gamma \end{bmatrix}.$$

Results are reported for $a_1 = -0.4$, $\gamma = 0.8$ and $\delta \in \{0, 0.2, 0.4\}$. It is easily seen that the true co-integrating rank in this model is $r_0 = 1$, and that although Assumption 1 is satisfied in all these cases, Assumption 2 of Swensen (2009) is only satisfied when $\delta = 0$. In particular, notice that Assumption 2 of Swensen (2009) requires that all of the eigenvalues of Γ_1 have modulus less than one. For $\delta = 0$ this is indeed the case. However, for $\delta = 0.2$ the largest eigenvalue of Γ_1 is 1.0 implying that the $I(1, r)$ conditions are not met when rank $r = 0$ is (wrongly) imposed; in fact the system will be integrated of order two. For $\delta = 0.4$ the largest eigenvalue is 1.2 and the $I(1, r)$ conditions are again not met, with the system becoming explosive.

All experiments were conducted using 5,000 replications. The number B of bootstrap replications used in the bootstrap algorithms was set to 399. All tests were conducted at the nominal 0.05 significance level. The VAR model was fitted with a restricted constant (i.e. deterministic case (ii) in Section 2), when calculating all of the tests for the reasons outlined in Nielsen and Rahbek (2000), relating to similarity with respect to the initial values. For the standard likelihood ratio tests asymptotic critical values as reported in Table 15.2 of Johansen (1996) were employed.

Table 1 reports the percentage of times that the asymptotic procedure of Johansen (1996) and Algorithm 1 of this paper with either i.i.d. or wild bootstrap re-sampling selects each of rank $r = 0, \dots, 4$. Also reported for each experiment is the percentage of times that step (iii) of Algorithm 2 of Swensen (2006) is failed; i.e., the percentage of times in practice (one would of course not know in practice if Assumption 2 of Swensen, 2009, held or not) that this algorithm terminates without providing an estimate of the co-integrating rank.

Tables 1 – 2 about here

For each of Models IID, GARCH and SV, where all of the proposed methods are asymptotically valid, we see that both versions of the bootstrap represent a considerable improvement over the asymptotic procedure of Johansen (1996). For Models IID and GARCH both bootstrap methods work very well, although the wild bootstrap based procedure is slightly more accurate than that based on the corresponding i.i.d. bootstrap for Model GARCH (consistent with what is reported in Cavaliere *et al.*, 2010a), with results very close to the asymptotic predictions from Theorem 1 even for $T = 100$. In contrast the asymptotic procedure can significantly over estimate the co-integrating rank in small samples. For example, for $T = 250$ under Model SV the asymptotic procedure selects rank one only about 75% of the time and rank two or more around 25% of the time. In contrast, the wild bootstrap procedure selects rank one about 94% of the time and rank two or more about 6% of the time. For Model SV the performance of the i.i.d. bootstrap is clearly inferior to that of the corresponding wild bootstrap; in the previous example it selects rank one about 85% of the time and rank two or more about 15% of the time.

For Model NSV only the procedure based on the wild bootstrap is asymptotically valid and this is clearly reflected in the small sample results. The asymptotic procedure is very badly behaved here and even for $T = 250$ is more likely to select rank two or more than the correct rank of one. The i.i.d. bootstrap is not quite so badly behaved but still selects rank equal to two or more with about 40% probability even for $T = 250$. The wild bootstrap does show some small sample distortions, perhaps not entirely surprisingly for a process which displays a twenty five fold increase in variance towards the end of the sample, but by $T = 250$ we see that it selects rank one about 90% of the time and rank two or more about 10% of the time.

It is interesting to consider the percentage of times that step (iii) of Algorithm 2 of Swensen (2006) is failed for each experiment. Under Models IID and GARCH step (iii) is never failed when $\delta = 0$, as one would hope. For Model SV it is almost never failed when $\delta = 0$. In contrast for Model NSV step (iii) is failed 29% and 13% of the time for $T = 100$ and $T = 250$, respectively, even though Assumption 2 of Swensen (2009) is met. For cases where $\delta > 0$ (such that Assumption 2 of Swensen, 2009, is failed and Algorithm 2 of Swensen or the corresponding wild bootstrap based procedure of Cavaliere *et al.*, 2010a and 2010b should not be used) we see that although step (iii) fails in the majority of cases when $\delta = 0.4$, this is not so when $\delta = 0.2$. For example, under Model IID with $\delta = 0.2$, even for $T = 250$ step (iii) is failed only 35% of the time. In other words 65% of the time step (iii) will be satisfied even though the algorithm should, in theory, have been terminated. Given this feature we certainly recommend that practitioners use the sequential bootstrap procedure outlined in Algorithm 1 of this paper.

For completeness, in Table 2 we conclude this section by reporting results corresponding to those in Table 1 for both the i.i.d. bootstrap Algorithm 2 of Swensen (2006) and the analogous wild bootstrap based procedure of Cavaliere *et al.* (2010a, 2010b) in those cases where step (iii) is (correctly) passed 100% of the time, such that a valid comparison may be made; i.e., for Models IID and GARCH when $\delta = 0$. Comparing these results with the corresponding results in Table 1, we see that even in these cases the procedures proposed in this paper display superior finite sample performance, in particular for $T = 100$, to the extant procedures. These findings are consistent with simulation results in Trenkler (2009) who finds that the finite sample properties of non-sequential bootstrap rank tests based on restricted estimation are superior to those based on unrestricted estimation.

6 Conclusions

In this paper we have developed a likelihood ratio-based bootstrap sequential procedure for determining the co-integrating rank of a system of $I(1)$ variables. In contrast to what has been established for extant bootstrap sequential procedures, our approach delivers consistent inference on the co-integration rank for general vector $I(1)$ processes, without the need for the additional assumptions outlined in Swensen (2009) to hold. Our approach can be used under both i.i.d. and wild bootstrap re-sampling schemes and consequently can allow for both conditional and unconditional heteroskedasticity in the underlying innovations. A small Monte Carlo experiment showed that our procedure works very well in finite samples, under both homoskedastic and heteroskedastic environments, and we recommend its use in practice.

A Appendix

PROOF OF LEMMA 1. For the proof it is convenient to introduce a normalization of the co-integrating vector of (1), which is the population counterpart of the normalization of the estimated co-integrating vectors given in (8). To do so, use that by Johansen (1996), the r_0 largest eigenvalues $(\hat{\lambda}_i)_{i=1,\dots,r_0}$ satisfy, as $T \rightarrow \infty$, the population eigenvalue problem,

$$|\lambda \Sigma_{\beta\beta} - \Sigma_{\beta 0} \Sigma_{00}^{-1} \Sigma_{0\beta}| = 0, \quad (\text{A.1})$$

with $\Sigma_{\beta\beta} := \text{Var}(\beta^{\#'} Z_{1t} | Z_{2t})$ and $\Sigma_{\beta 0} := \text{Cov}(\beta^{\#'} Z_{1t}, Z_{0t} | Z_{2t})$. Let $\kappa := (\kappa_1, \dots, \kappa_{r_0})$ denote eigenvectors corresponding to the eigenvalues $\lambda_1 > \lambda_2 > \dots > \lambda_{r_0} > 0$, such that $\kappa' \Sigma_{\beta\beta} \kappa = I_{r_0}$. We can then define $\beta_0^{\#} := \beta^{\#} \kappa$ ($\beta_0 := \beta \kappa$) and $\alpha_0 := \alpha (\kappa')^{-1}$. Observe that, $\alpha \beta^{\#'} = \alpha_0 \beta_0^{\#'}$, while also

$$\Sigma_{\beta_0 \beta_0} = I_{r_0} \quad \text{and} \quad \Sigma_{\beta_0 0} \Sigma_{00}^{-1} \Sigma_{0 \beta_0} = \text{diag}(\lambda_1, \dots, \lambda_{r_0}), \quad (\text{A.2})$$

with $\Sigma_{\beta_0 \beta_0} := \text{Var}(\beta_0^{\#'} Z_{1t} | Z_{2t})$ and $\Sigma_{\beta_0 0} := \text{Cov}(\beta_0^{\#'} Z_{1t}, Z_{0t} | Z_{2t})$. Indeed, the relations in (A.2) are the population equivalents of the sample normalisations in (8). The vector of unique true parameters is denoted by $\theta_0 := (\alpha_0, \beta_0, \Psi_0, \Omega_0, \mu_0)$, and $\beta_0^{\#} = (\beta_0', \rho_0')$, cf. (1) and (2).

Consider first the case of *no deterministics*; that is, $\mu_0 = 0$ and hence $\beta_0 = \beta_0^{\#}$. With $\hat{\beta} = \hat{\beta}^{(r_0)}$, the MLE under the true rank r_0 , then as in Johansen (1995, proof of Lemma

13.1), $\bar{\beta}'_0(\hat{\beta} - \beta_0) \xrightarrow{p} 0$ and $T^{1/2}\beta'_{0\perp}(\hat{\beta} - \beta_0) \xrightarrow{p} 0$, using continuity of the eigenvectors and eigenvalues, as the latter are distinct. Therefore, as $\hat{\beta}^{(r)} = \hat{v}K_p^{(r)} = \hat{\beta}K_{r_0}^{(r)}$, it follows directly that,

$$\bar{\beta}'_0(\hat{\beta}^{(r)} - \beta_0^{(r)}) \xrightarrow{p} 0 \quad \text{and} \quad T^{1/2}\beta'_{0\perp}(\hat{\beta}^{(r)} - \beta_0^{(r)}) \xrightarrow{p} 0 \quad (\text{A.3})$$

where $\beta_0^{(r)} := \beta_0 K_{r_0}^{(r)}$. Next, consider the QMLE $\hat{\alpha}^{(r)}$:

$$\hat{\alpha}^{(r)} = S_{01}\hat{\beta}^{(r)} = S_{01} \left(\beta_0 \bar{\beta}'_0 + \beta_{0\perp} \bar{\beta}'_{0\perp} \right) \hat{\beta}^{(r)} \xrightarrow{p} \Sigma_{0\beta_0} K_{r_0}^{(r)} = \alpha_0 K_{r_0}^{(r)} =: \alpha_0^{(r)}. \quad (\text{A.4})$$

Also, with $\Upsilon_{02} = \text{Cov}(Z_{0t}, Z_{2t})$, $\Upsilon_{\beta_0 2} = \text{Cov}(\beta'_0 Z_{1t}, Z_{2t})$ and $\Upsilon_{22} = \text{Var}(Z_{2t})$,

$$\begin{aligned} \hat{\Psi}^{(r)} &= \left(M_{02} - \hat{\alpha}^{(r)} \hat{\beta}^{(r)'} M_{12} \right) M_{22}^{-1} \xrightarrow{p} \Psi_0^{(r)} := \left(\Upsilon_{02} - \alpha_0 K_{r_0}^{(r)} K_{r_0}^{(r)'} \Upsilon_{\beta_0 2} \right) \Upsilon_{22}^{-1} \\ &= \Psi_0 + \alpha_0 K_{r_0, \perp}^{(r)} K_{r_0, \perp}^{(r)'} \Upsilon_{\beta_0 2} \Upsilon_{22}^{-1} \end{aligned} \quad (\text{A.5})$$

with $K_{r_0, \perp}^{(r)} = (0, I_{r_0-r})'$. Likewise,

$$\hat{\Omega}^{(r)} = S_{00} - \hat{\alpha}^{(r)} \hat{\alpha}^{(r)'} \xrightarrow{p} \Omega_0^{(r)} := \Sigma_{00} - \alpha_0 K_{r_0}^{(r)} K_{r_0}^{(r)'} \alpha_0' = \Omega_0 + \alpha_0 K_{r_0, \perp}^{(r)} K_{r_0, \perp}^{(r)'} \alpha_0' > 0. \quad (\text{A.6})$$

To see that the characteristic polynomial satisfies the I(1, r) conditions under smaller rank $r \leq r_0$, first rewrite the DGP as,

$$Z_{0t} = \alpha_0^{(r)} \beta_0^{(r)'} Z_{1t} + \Psi_0^{(r)} Z_{2t} + \varepsilon_t^{(r)}, \quad (\text{A.7})$$

where the innovations $\varepsilon_t^{(r)}$ are given by,

$$\varepsilon_t^{(r)} = \varepsilon_t + \alpha_0 K_{r_0, \perp}^{(r)} K_{r_0, \perp}^{(r)'} (\beta_0' Z_{1t} - \Upsilon_{\beta_0 2} \Upsilon_{22}^{-1} Z_{2t}). \quad (\text{A.8})$$

Observe that $\beta_0^{(r)'} Z_{1t}$ and Z_{2t} in (A.7) are uncorrelated with the innovation term $\varepsilon_t^{(r)}$, which is used below to establish the algebraic I(1, r) results for the characteristic polynomial $A^{(r)}(z)$ in (10). With $\mathbb{X}_t := (X_t', \dots, X_{t-k+1}')'$ the system can be written in companion form as,

$$\Delta \mathbb{X}_t = \mathbb{A}^{(r)} \mathbb{B}^{(r)'} \mathbb{X}_{t-1} + \mathbb{E}_t^{(r)} \quad (\text{A.9})$$

with $\mathbb{E}_t^{(r)} := (\varepsilon_t^{(r)'}, 0, \dots, 0)'$, \mathbb{X}_0 fixed and with $\Psi_0^{(r)} = (\Gamma_{1,0}^{(r)}, \dots, \Gamma_{k-1,0}^{(r)})$,

$$\begin{aligned} \mathbb{A}^{(r)} &:= \begin{pmatrix} \alpha_0^{(r)} & \Psi_0^{(r)} \\ 0 & I_{p(k-1)} \end{pmatrix} = \begin{pmatrix} \alpha_0^{(r)} & \Gamma_{1,0}^{(r)} & \cdots & \cdots & \Gamma_{k-1,0}^{(r)} \\ 0 & I_p & 0 & \cdots & 0 \\ 0 & 0 & I_p & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & I_p \end{pmatrix} \\ \mathbb{B}^{(r)} &:= \begin{pmatrix} \beta_0^{(r)} & I_p & 0 & \cdots & 0 \\ 0 & -I_p & I_p & \cdots & 0 \\ 0 & 0 & -I_p & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & -I_p \end{pmatrix}. \end{aligned} \quad (\text{A.10})$$

Observe first, that by Assumption 1, $\mathbb{Y}_t := \mathbb{B}^{(r)'} \mathbb{X}_t$ is covariance stationary with covariance $\Sigma_{YY} > 0$, which furthermore from (A.9), using the mentioned uncorrelatedness², is the solution to,

$$\Sigma_{YY} = \Phi^{(r)} \Sigma_{YY} \Phi^{(r)'} + \Sigma_{EE}, \quad (\text{A.11})$$

where $\Phi^{(r)} = (I_{r+p(k-1)} + \mathbb{B}^{(r)'} \mathbb{A}^{(r)})$ and $\Sigma_{EE} = \text{Var}(\mathbb{B}^{(r)'} \mathbb{E}_t^{(r)})$. From the definition of $\varepsilon_t^{(r)}$ in (A.8), $\text{Var}(\varepsilon_t^{(r)}) > 0$ and, as $\mathbb{B}^{(r)'} \mathbb{E}_t^{(r)} = (\varepsilon_t^{(r)'} \beta_0^{(r)}, \varepsilon_t^{(r)'} 0, \dots, 0)'$ it follows that $\Sigma_{EE} \geq 0$, with $V' \Sigma_{EE} V = 0$, and $V_{\perp}' \Sigma_{EE} V_{\perp} > 0$, where

$$V = \begin{pmatrix} I_r & 0 & \cdots & 0 \\ -\beta_0^{(r)} & 0 & & \vdots \\ 0 & I_p & & \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & I_p \end{pmatrix}, \quad V_{\perp} = \begin{pmatrix} \beta^{(r)'} \\ I_p \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \quad (\text{A.12})$$

As $\Phi^{(r)}$ solves (A.11), and since $\Sigma_{YY} > 0$ and $\Sigma_{EE} \geq 0$, the spectral radius of $\Phi^{(r)}$ satisfies $\rho(\Phi^{(r)}) \leq 1$. Suppose, for λ an eigenvalue of $\Phi^{(r)'}'$, $|\lambda| = 1$. Then, using (A.11), the corresponding eigenvector v is in the space spanned by V , $v \in \text{col}(V)$. However, as

$$\begin{aligned} \Phi^{(r)'} V &= (I_{r+p(k-1)} + \mathbb{A}^{(r)'} \mathbb{B}^{(r)}) V \\ &= \begin{pmatrix} I_r & 0 & \cdots & 0 \\ 0 & I_p & & \vdots \\ 0 & 0 & \ddots & \\ \vdots & & \ddots & I_p \\ 0 & \cdots & & 0 \end{pmatrix}, \end{aligned}$$

any $v \in \text{col}(V)$ is not an eigenvector of $\Phi^{(r)'}'$. Thus, we conclude $\rho(\Phi^{(r)}) < 1$. Finally, consider the eigenvalue problem,

$$\left| (1 - z) I_{pk} - \mathbb{A}^{(r)} \mathbb{B}^{(r)'} z \right| = 0. \quad (\text{A.13})$$

Observe that $\mathbb{B}^{(r)'} \mathbb{A}^{(r)} \neq 0$ such that $N := (\mathbb{B}^{(r)}, \mathbb{A}_{\perp}^{(r)})$ has full rank, where

$$\mathbb{A}_{\perp}^{(r)} = (I_p, -\Gamma_{1,0}^{(r)}, \dots, -\Gamma_{k-1,0}^{(r)})' \alpha_{0,\perp}^{(r)}, \quad \mathbb{B}_{\perp}^{(r)} = (I_p, \dots, I_p)' \beta_{0,\perp}^{(r)}.$$

Multiplying from the left and right by N and its transpose respectively in (A.13), shows that there are $(p - r)$ roots at $z = 1$, while the remaining satisfy $|z| > 1$ since $\rho(\Phi^{(r)}) < 1$. This finishes the proof of Lemma 1 for the case of no deterministic terms.

Consider next the case of a *restricted constant*; that is, $\mu_0 = \alpha_0 \rho_0'$ and $\beta_0^{\#} = (\beta_0', \rho_0')'$. Define $\gamma_0^{\#} := (\beta_{0,\perp}', 0)'$ and $\xi_0^{\#} := (-\rho_0 \bar{\beta}_0', 1)'$ such that $(\beta_0^{\#}, \gamma_0^{\#}, \xi_0^{\#})$ spans $\mathcal{R}^{p^{\#}}$. As before, it holds that

$$\left((\hat{\beta}^{\#} - \beta_0^{\#})' \bar{\beta}_0^{\#}, T^{1/2} (\hat{\beta}^{\#} - \beta_0^{\#})' \bar{\gamma}_0^{\#}, (\hat{\beta}^{\#} - \beta_0^{\#})' \bar{\xi}_0^{\#} \right) \xrightarrow{p} 0.$$

²This is the vital difference between our Algorithm 1 and Algorithm 2 of Swensen (2006); in the latter, $\Psi_0^{(r)}$ is replaced by Ψ_0 , this resulting from the use of the unrestricted estimator of Ψ_0 .

Therefore, as $\hat{\beta}^{\#(r)} = \hat{v}K_{p^\#}^{(r)} = \hat{\beta}^\# K_{r_0}^{(r)}$, we have the following results:

$$\bar{\beta}_0^{\#'}(\hat{\beta}^{\#(r)} - \beta_0^{\#(r)}) \xrightarrow{p} 0, \quad T^{1/2}\gamma_0^{\#'}(\hat{\beta}^{\#(r)} - \beta_0^{\#(r)}) \xrightarrow{p} 0 \quad \text{and} \quad \xi_0^{\#'}(\hat{\beta}^{\#(r)} - \beta_0^{\#(r)}) \xrightarrow{p} 0.$$

The results follow proceeding as in the case of no deterministics. Specifically, in (A.4)-(A.8) replace β_0 by $\beta_0^\#$, while from (A.9) and onwards, the term $\mathbb{A}^{(r)}\mathbb{R}^{(r)'} should be added to the right hand side of (A.9), where$

$$\mathbb{R}^{(r)} := \left(\rho_0^{(r)}, 0, \dots, 0 \right). \quad (\text{A.14})$$

Consider finally the case of a *restricted linear trend*; that is, $\mu_0 = \mu_{10} + \alpha_0 \rho_0'$ and $\beta_0^\# = (\beta_0', \rho_0')'$. For the asymptotic analysis in this case, Rahbek *et al.* (1999, proof of Theorem 4.2) apply the non-orthogonal basis $(\beta_0^\#, \gamma_0^\#, \xi_0^\#)$ for $\mathcal{R}^{p^\#}$, where $\gamma_0^\# := (\beta_{0\perp}', -\gamma_0')'$ and $\xi_0^\# := (0, 1)'$, with $\gamma_0 := \beta_{0\perp}' C_0 (\mu_{10} + \Gamma_0 \bar{\beta}_0 \rho_0')$ and $C_0 := \beta_{0\perp} (\alpha_{0\perp}' \Gamma_0 \beta_{0\perp})^{-1} \alpha_{0\perp}'$. With, $\hat{\beta}^\# - \beta_0^\# := \beta_0^\# \hat{b} + \gamma_0^\# \hat{b}_\gamma + \xi_0^\# \hat{b}_\xi$, then by Rahbek *et al.* (1999), $\hat{b} = o_p(1)$, $\hat{b}_\gamma = o_p(T^{1/2})$, $\hat{b}_\xi = o_p(T)$, such that $\hat{\beta}^{\#(r)} = \hat{\beta}^\# K_{r_0}^{(r)}$ converges as desired. Proceed as in the case of a restricted constant, replacing β_0 by $\beta_0^\#$ in (A.4)-(A.8), and adding the term $\mathbb{A}^{(r)}\mathbb{R}^{(r)'}t + \mathbb{M}_1^{(r)}$ in (A.9), with $\mathbb{R}^{(r)}$ defined in (A.14) and $\mathbb{M}_1^{(r)} := (\mu_{10}^{(r)'}, 0, \dots, 0)'$. Last, define the covariance stationary process \mathbb{Y}_t in this case as $\mathbb{Y}_t := \mathbb{B}^{(r)'}\mathbb{X}_t + \mathbb{R}^{(r)'}t$. \square

PROOF OF PROPOSITION 1: For $r = r_0$ the result is established in Lemma A.4 of Cavaliere *et al.* (2010a). Now consider the case where $r < r_0$. Set $\mathbb{X}_t^* := (X_t^{*'}, \dots, X_{t-k+1}^{*'})'$ and $\mathbb{X}_0^* := 0$, and use the companion form in (A.9) in the proof of Lemma 1 with $\varepsilon_t^{(r)}$ replaced by $\varepsilon_{r,t}^*$, to see directly that $X_t^* = (I_p, 0, \dots, 0) \mathbb{X}_t^*$ has the representation,

$$X_{r,t}^* = \hat{C}^{(r)} \sum_{i=1}^t \varepsilon_{r,t}^* + S_{r,t} T^{1/2}$$

where $\hat{C}^{(r)} = \hat{\beta}_\perp^{(r)} \left(\hat{\alpha}_\perp^{(r)'} \hat{\Gamma}^{(r)} \hat{\beta}_\perp^{(r)} \right)^{-1} \hat{\alpha}_\perp^{(r)'}$, $\hat{\Gamma}^{(r)} = \sum_{i=1}^{k-1} \hat{\Gamma}_i^{(r)} - I$, and,

$$S_{r,t} = (\hat{\alpha}^{(r)}, \hat{\Psi}^{(r)}) (\hat{\mathbb{B}}' \hat{\mathbb{A}})^{-1} \sum_{i=0}^{t-1} \hat{\Phi}^i (T^{-1/2} \hat{\mathbb{B}}' e_{r,t-i}^*).$$

Here $\hat{\Phi} := (I_{r+p(k-1)} + \hat{\mathbb{B}}' \hat{\mathbb{A}})$ with the matrices $\hat{\mathbb{A}}$ and $\hat{\mathbb{B}}$ defined as $\mathbb{A}^{(r)}$ and $\mathbb{B}^{(r)}$ of (A.10) with $\alpha_0^{(r)}$ and $\beta_0^{(r)}$ replaced by the corresponding estimators $\hat{\alpha}^{(r)}$, $\hat{\beta}^{(r)}$, and $e_{r,t}^* := (\varepsilon_{r,t}^{*'}, 0, \dots, 0)'$. Next, note that

$$\max_{t=1, \dots, T} \|S_{r,t}\| \leq \max_{t=1, \dots, T} \left\| (\hat{\alpha}^{(r)}, \hat{\Psi}^{(r)}) (\hat{\mathbb{B}}' \hat{\mathbb{A}})^{-1} \sum_{i=0}^{t-1} \hat{\Phi}^i \left(T^{-1/2} \hat{\mathbb{B}}' e_{r,t-i}^* \right) \right\| \leq \psi_T \max_{t=1, \dots, T} \|T^{-1/2} \eta_t^*\|$$

where $\eta_t^* = \hat{\mathbb{B}}' e_{r,t}^* = (\hat{\beta}^{(r)}, I_p, 0, \dots, 0)' \varepsilon_{r,t}^*$ and $\psi_T = \|(\hat{\alpha}^{(r)}, \hat{\Psi}^{(r)}) (\hat{\mathbb{B}}' \hat{\mathbb{A}})^{-1}\| \left\| \sum_{i=0}^{T-1} \hat{\Phi}^i \right\|$. It follows that $\psi_T \xrightarrow{p} \psi = \|(\alpha_0^{(r)}, \Psi_0^{(r)}) (\mathbb{B}^{(r)'} \mathbb{A}^{(r)})^{-1}\| \|(\mathbb{B}^{(r)'} \mathbb{A}^{(r)})^{-1}\|$ by using the established

consistency of the estimators Lemma 1 and the definition of $\hat{\Phi}$. In particular, note that for sufficiently large T we have, by continuity, that $\rho(\hat{\Phi}) < 1$, which implies that $\|\hat{\Phi}^i\| \leq \text{const.} \lambda^i$ for some $0 < \lambda < 1$, uniformly over i . Finally, showing that $P^*(\max_{t=1,\dots,T} \|T^{-1/2} \eta_t^*\| > \eta)$ is of order $o_p(1)$ implies the desired result that $P^*(\max_{t=1,\dots,T} \|S_{r,t}\| > \eta) \xrightarrow{p} 0$. This holds by showing $P^*(T^{-1/2} \max_{t=1,\dots,T} \|\varepsilon_{r,t}^*\| > \eta) = o_p(1)$. For the wild bootstrap, observe that with $\bar{\varepsilon}_r := T^{-1} \sum_{t=1}^T \hat{\varepsilon}_{r,t}$,

$$E^*(\varepsilon_{r,t}^{*'} \varepsilon_{r,t}^*)^2 = [(\hat{\varepsilon}_{r,t} - \bar{\varepsilon}_r)' (\hat{\varepsilon}_{r,t} - \bar{\varepsilon}_r)]^2 = (\hat{\varepsilon}_{r,t}' \hat{\varepsilon}_{r,t} - \bar{\varepsilon}_r' \bar{\varepsilon}_r)^2,$$

while for the i.i.d. bootstrap,

$$E^*(\varepsilon_{r,t}^{*'} \varepsilon_{r,t}^*)^2 = \frac{1}{T} \sum_{t=1}^T (\hat{\varepsilon}_{r,t}' \hat{\varepsilon}_{r,t} - \bar{\varepsilon}_r' \bar{\varepsilon}_r)^2.$$

Thus in both cases one has, by Chebychev's inequality,

$$\begin{aligned} P^*\left(T^{-1/2} \max_{t=1,\dots,T} \|\varepsilon_{r,t}^*\| > \eta\right) &\leq \frac{1}{\eta^4 T^2} \sum_{t=1}^T E^*(\varepsilon_{r,t}^{*'} \varepsilon_{r,t}^*)^2 \\ &= \frac{3}{\eta^4 T^2} \sum_{t=1}^T (\hat{\varepsilon}_{r,t}' (\hat{\varepsilon}_{r,t} - \bar{\varepsilon}_r))^2 \xrightarrow{p} 0 \end{aligned} \quad (\text{A.15})$$

since $T^{-1} \sum_{t=1}^T (\hat{\varepsilon}_{r,t}' \hat{\varepsilon}_{r,t})^2 = O_p(1)$ under the assumption that ε_t has bounded fourth moment. To see this note that by definition, cf. (A.8),

$$\begin{aligned} \hat{\varepsilon}_{r,t} &= Z_{0t} - \hat{\alpha}^{(r)} \hat{\beta}^{\#(r)'} Z_{1t} - \hat{\Psi}^{(r)} Z_{2t} \\ &= \varepsilon_t - \left(\hat{\alpha}^{(r)} \hat{\beta}^{\#(r)'} - \alpha_0 \beta_0^{\#'} \right) Z_{1t} - \left(\hat{\Psi}^{(r)} - \Psi_0 \right) Z_{2t}, \end{aligned}$$

where we recall that $\beta_0^{\#} = \beta_0$ in the case of no deterministics. Next, observe first that,

$$\begin{aligned} \hat{\alpha}^{(r)} \hat{\beta}^{\#(r)'} - \alpha_0 \beta_0^{\#'} &= \left(\hat{\alpha} K_{r_0}^{(r)} K_{r_0}^{(r)'} \hat{\beta}^{\#'} - \alpha_0 K_{r_0}^{(r)} K_{r_0}^{(r)'} \beta_0^{\#'} \right) - \alpha_0 K_{r_0,\perp}^{(r)} K_{r_0,\perp}^{(r)'} \beta_0^{\#'} \\ &= (\hat{\alpha} - \alpha_0) K_{r_0}^{(r)} K_{r_0}^{(r)'} \beta_0^{\#'} - \hat{\alpha} K_{r_0}^{(r)} K_{r_0}^{(r)'} \left(\hat{\beta}^{\#} - \beta_0^{\#} \right)' - \alpha_0 K_{r_0,\perp}^{(r)} K_{r_0,\perp}^{(r)'} \beta_0^{\#'} \end{aligned} \quad (\text{A.16})$$

Likewise,

$$\begin{aligned} \hat{\Psi}^{(r)} - \Psi_0 &= \left(M_{02} - \hat{\alpha}^{(r)} \hat{\beta}^{\#(r)'} M_{12} \right) M_{22}^{-1} - (\Upsilon_{02} - \alpha_0 \Upsilon_{\beta_0 2}) \Upsilon_{22}^{-1} \\ &= (M_{02} M_{22}^{-1} - \Upsilon_{02} \Upsilon_{22}^{-1}) - \left(\hat{\alpha}^{(r)} \hat{\beta}^{\#(r)'} - \alpha_0 \beta_0^{\#'} \right) M_{12} M_{22}^{-1} \\ &\quad - \alpha_0 \left(\beta_0^{\#'} M_{12} M_{22}^{-1} - \Upsilon_{\beta_0 2} \Upsilon_{22}^{-1} \right). \end{aligned} \quad (\text{A.17})$$

Using (A.16) and (A.17), and collecting terms, it follows that,

$$\hat{\varepsilon}_{r,t} = \varepsilon_{r,t} + v_{r,t} \quad (\text{A.18})$$

where $\varepsilon_{r,t}$ is given in (A.8), while

$$v_{r,t} = \left[\hat{\alpha} K_{r_0}^{(r)} K_{r_0}^{(r)'} \left(\hat{\beta}^\# - \beta_0^\# \right)' - (\hat{\alpha} - \alpha_0) K_{r_0}^{(r)} K_{r_0}^{(r)'} \beta_0^{\#'} \right] R_{1t} \\ + \left[\alpha_0 \left(\beta_0^{\#'} M_{12} M_{22}^{-1} - \Upsilon_{\beta_0 2} \Upsilon_{22}^{-1} \right) - (M_{02} M_{22}^{-1} - \Upsilon_{02} \Upsilon_{22}^{-1}) \right] Z_{2t}$$

with $R_{1t} = Z_{1t} - M_{12} M_{22}^{-1} Z_{2t}$, that is, Z_{1t} corrected for Z_{2t} by OLS regression. Thus, as claimed above,

$$\frac{1}{T} \sum_{t=1}^T (\hat{\varepsilon}_{r,t}' \hat{\varepsilon}_{r,t})^2 = \frac{1}{T} \sum_{t=1}^T (\varepsilon_t' \varepsilon_t)^2 + o_p(1), \quad (\text{A.19})$$

by using standard arguments for the stationary processes ε_t , $\beta_0^{\#'} Z_{1t}$ and Z_{2t} in combination with consistency of the parameters in the definition of $v_{r,t}$, while for (cross) products in terms of the non-stationary process $\gamma_0^{\#'} Z_{1t}$ (and $\xi_0^{\#'} Z_{1t}$ in the case of a restricted trend), standard arguments for non-stationary processes in combination with super consistency give the desired result. We have here used the notation introduced in the proof of Lemma 1, such that the vectors in the basis $(\beta_0^\#, \gamma_0^\#, \xi_0^\#)$, which are defined for the case of a restricted constant and linear trend respectively, in the case of no deterministic terms are set equal to $(\beta_0, \bar{\beta}_{0\perp})$.

Next, $T^{-1/2} X_{r, [Tu]}^* \xrightarrow{w_p} C_0^{(r)} W(u)$, follows by establishing, $T^{-1/2} \sum_{t=1}^{[Tu]} \varepsilon_{r,t}^* \xrightarrow{w_p} W(\cdot)$, which for the wild bootstrap, as in Cavaliere *et al.* (2010a), is implied by the pointwise convergence,

$$\frac{1}{T} \sum_{t=1}^{[Tu]} \hat{\varepsilon}_{r,t} \hat{\varepsilon}_{r,t}' = \frac{1}{T} \sum_{t=1}^{[Tu]} \varepsilon_{r,t} \varepsilon_{r,t}' + o_p(1) \xrightarrow{p} u \Omega_0^{(r)}, \quad (\text{A.20})$$

see also (A.19). For the i.i.d. bootstrap, the result follows as in Swensen (2006, Lemma S2) using (A.20) and finite fourth order moments of $\varepsilon_{r,t}$. \square

PROOF OF PROPOSITION 2: As in the proof of Theorem 3 of Cavaliere *et al.* (2010a), this follows immediately by the results in Proposition 1 using standard arguments and defining

$$B_{p-r} := \left(\alpha_\perp^{(r)'} \Omega^{(r)} \alpha_\perp^{(r)} \right)^{-1/2} \alpha_\perp^{(r)'} W. \quad \square$$

PROOF OF THEOREM 1: The proof mimics that of Theorem 12.3 in Johansen (1996). In brief, note first that, as established in Proposition 2, for each rank $r < r_0$ being tested, the bootstrap trace statistic for rank r has the usual asymptotic null distribution, $\text{tr}(Q_{r,\infty})$. But since the original trace statistics $Q_{r,T}$ diverge when an incorrect rank $r < r_0$ is imposed, see (5), our sequential algorithm will consequently reject any $r < r_0$ with probability one in the limit. Finally, since (11) also holds for $r = r_0$, for a chosen significance level η , the overall probability of selecting rank r_0 will therefore equal $1 - \eta \cdot \mathbb{I}(r_0 < p)$, in the limit. \square

References

Andrews, D.W.K., and M. Buchinsky (2001), Evaluation of a three-step method for choosing the number of bootstrap repetitions, *Journal of Econometrics* 103, 345-386.

- Cavaliere G. and A.M.R. Taylor (2008), Bootstrap unit root tests for time series with non-stationary volatility, *Econometric Theory* 24, 43-71.
- Cavaliere, G., A. Rahbek and A.M.R. Taylor (2010a), Co-integration rank testing under conditional heteroskedasticity, forthcoming in *Econometric Theory*.
- Cavaliere, G., A. Rahbek and A.M.R. Taylor (2010b), Testing for co-integration in vector autoregressions with non-stationary volatility, Granger Center Discussion Paper 07/02, forthcoming in *Journal of Econometrics*.
- Engle, R.F. and C.W.J. Granger (1987), Co-integration and error-correction: representation, estimation and testing, *Econometrica* 55, 251—276.
- Giné, E. and J. Zinn (1990), Bootstrapping general empirical measures, *Annals of Probability* 18, 851–869.
- Gonçalves, S. and L. Kilian (2004), Bootstrapping autoregressions with conditional heteroskedasticity of unknown form, *Journal of Econometrics* 123, 89-120.
- Hansen, B.E. (1996), Inference when a nuisance parameter is not identified under the null hypothesis, *Econometrica* 64, 413–430.
- Harris, R.I.D. and G. Judge (1998), Small sample testing for cointegration using the bootstrap approach, *Economics Letters* 58, 31–37.
- Johansen, S. (1996), *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford: Oxford University Press.
- Johansen, S. (2002), A small sample correction of the test for cointegrating rank in the vector autoregressive model, *Econometrica* 70, 1929–1961.
- Liu, R.Y. (1988), Bootstrap procedure under some non-iid models, *Annals of Statistics* 16, 1696-1708.
- Mammen, E. (1993), Bootstrap and wild bootstrap for high dimensional linear models, *Annals of Statistics* 21, 255–285.
- Mantalos, P. and G. Shukur (2001), Bootstrapped Johansen tests for cointegration relationships: a graphical analysis, *Journal of Statistical Computation and Simulation* 68, 351–371.
- Nielsen, B. and A. Rahbek (2000), Similarity issues in cointegration analysis, *Oxford Bulletin of Economics and Statistics* 62, 5-22.
- Paruolo, P. (2001), The power of lambda max, *Oxford Bulletin of Economics and Statistics* 63, 395-403.
- Rahbek, A., H.C. Kongsted and C. Jørgensen (1999), Trend stationarity in the I(2) cointegration model, *Journal of Econometrics* 90, 265-289.
- Swensen, A.R. (2006), Bootstrap algorithms for testing and determining the cointegration rank in VAR models, *Econometrica* 74, 1699–1714.

- Swensen, A.R. (2009), Corrigendum to "Bootstrap algorithms for testing and determining the cointegration rank in VAR models", *Econometrica* 77, 1703–1704.
- Trenkler, C. (2009), Bootstrapping systems cointegration tests with a prior adjustment for deterministic terms, *Econometric Theory* 25, 243-269.
- van Giersbergen, N.P.A. (1996), Bootstrapping the Trace Statistics in VAR Models: Monte Carlo Results and Applications, *Oxford Bulletin of Economics and Statistics* 58, 391–408.
- Wu, C.F.J. (1986), Jackknife, bootstrap, and other resampling methods, *Annals of Statistics* 14, 1261–1295.

Table 1: Asymptotic and Bootstrap Sequential Procedures for Selecting the Co-integration Rank. True Rank is 1.

Model	δ	T	Fail %	ASYMPTOTIC TEST					I.I.D. BOOTSTRAP					WILD BOOTSTRAP				
				$r = 0$	1	2	3	4	$r = 0$	1	2	3	4	$r = 0$	1	2	3	4
IID	0	100	0%	0.0	75.2	20.6	3.6	0.6	0.6	93.5	4.9	0.7	0.3	0.7	94.1	4.3	0.6	0.3
		250	0%	0.0	88.7	9.9	1.1	0.2	0.0	95.0	4.4	0.5	0.2	0.0	95.1	4.1	0.5	0.3
	0.2	100	29%	0.0	75.7	20.4	3.3	0.5	0.9	93.1	5.0	0.8	0.3	1.1	93.8	4.2	0.6	0.2
		250	35%	0.0	89.0	9.6	1.2	0.2	0.0	95.3	4.1	0.5	0.2	0.0	95.4	3.9	0.5	0.2
	0.4	100	95%	0.0	77.5	18.7	3.2	0.6	0.7	93.9	4.5	0.8	0.2	0.9	94.2	4.1	0.7	0.1
		250	100%	0.0	89.7	9.1	1.0	0.2	0.0	95.1	4.4	0.4	0.1	0.0	95.4	4.1	0.4	0.2
GARCH	0	100	0%	0.0	73.5	22.0	3.6	0.8	1.2	92.6	5.2	0.6	0.4	2.1	92.9	4.3	0.4	0.3
		250	0%	0.0	86.6	11.9	1.3	0.1	0.0	93.8	5.5	0.6	0.1	0.0	94.9	4.5	0.4	0.2
	0.2	100	28%	0.0	74.4	21.3	3.6	0.7	1.2	92.4	5.4	0.6	0.3	2.5	92.5	4.2	0.5	0.3
		250	35%	0.0	87.4	11.2	1.3	0.2	0.0	93.7	5.5	0.6	0.2	0.0	94.8	4.5	0.5	0.2
	0.4	100	93%	0.2	75.6	20.4	3.2	0.6	2.3	91.8	5.1	0.5	0.4	3.8	91.3	4.3	0.4	0.2
		250	100%	0.0	87.6	11.1	1.2	0.1	0.0	93.6	5.5	0.6	0.2	0.0	94.9	4.3	0.6	0.2
SV	0	100	2%	0.3	63.7	28.6	6.3	1.1	2.8	82.2	12.6	1.7	0.7	8.7	85.2	5.0	0.8	0.3
		250	1%	0.0	75.4	20.8	3.3	0.5	0.0	84.1	13.5	1.9	0.5	0.1	94.1	4.7	0.8	0.3
	0.2	100	35%	1.2	63.8	28.1	5.9	1.0	5.5	80.6	11.3	1.7	0.8	12.9	81.5	4.4	0.9	0.3
		250	37%	0.0	75.7	20.7	3.2	0.4	0.0	84.5	13.2	1.8	0.5	0.2	93.9	4.8	0.7	0.4
	0.4	100	87%	2.4	65.3	26.5	5.0	0.8	10.2	76.4	11.1	1.6	0.7	19.7	74.9	4.2	0.8	0.3
		250	98%	0.0	76.3	20.3	2.9	0.4	0.1	84.8	12.8	1.9	0.5	0.7	93.3	4.9	0.7	0.4
NSV	0	100	29%	0.4	55.2	36.4	7.0	0.9	3.0	76.3	18.6	1.9	0.2	13.3	75.6	9.5	1.3	0.2
		250	13%	0.0	48.1	43.9	7.2	0.8	0.0	61.4	34.7	3.5	0.4	0.0	90.9	8.3	0.7	0.1
	0.2	100	45%	0.3	56.7	34.7	7.2	1.0	3.1	76.4	18.6	1.6	0.2	14.9	74.5	9.2	1.2	0.2
		250	42%	0.0	48.6	43.5	7.1	0.8	0.0	61.0	35.4	3.2	0.4	0.0	90.4	8.6	0.9	0.1
	0.4	100	75%	0.2	58.4	34.3	6.4	0.7	3.1	78.2	17.1	1.3	0.2	13.8	76.5	8.5	1.0	0.2
		250	92%	0.0	47.7	44.3	7.3	0.7	0.0	59.2	36.6	3.8	0.4	0.0	90.4	8.6	0.9	0.1

Note: The column headed ‘Fail %’ reports the percentage of times that step (iii) of Algorithm 2 of Swensen (2006) is failed.

Table 2: Unrestricted Bootstrap Sequential Procedures for Selecting the Co-integration Rank. True Rank is 1.

Model	T	δ	I.I.D. BOOTSTRAP					WILD BOOTSTRAP				
			$r = 0$	1	2	3	4	$r = 0$	1	2	3	4
IID	100	0	0.8	91.2	6.5	1.1	0.5	0.6	91.0	6.8	1.1	0.4
	250	0	0.0	94.5	4.7	0.5	0.3	0.0	94.3	4.9	0.6	0.2
GARCH	100	0	2.2	90.1	6.3	0.9	0.6	1.2	89.9	7.1	1.0	0.8
	250	0	0.0	94.2	5.0	0.5	0.2	0.0	93.1	6.0	0.8	0.2